Bayesian Hierarchical Model Characterization of Model Error in Ocean Data Assimilation and Forecasts

Ralph F. Milliff
Colorado Research Associates Division, NWRA
3380 Mitchell Lane
Boulder, CO 80301
phone: (303) 415-9701 fax: (303) 415-9702 email: milliff@cora.nwra.com

Christopher K. Wikle
Department of Statistics, University of Missouri
146 Middlebush
Columbia, MO 65211
phone: (573) 882-9659 fax: (573) 884-5524 email: wikle@stat.missouri.edu

L. Mark Berliner and Radu Herbei
Department of Statistics, The Ohio State University
1958 Neil Ave.
Columbus, OH 43210
phone: (614) 292-0291 fax: (614) 292-2096 email: mb@stat.osu.edu

Award Number: N00014-10-C-0354

LONG-TERM GOALS

We seek to focus quantitative uncertainty management attributes of the Bayesian Hierarchical Model (BHM) methodology on the identification, characterization, and evolution of irreducible model error in ocean data assimilation and forecast systems.

OBJECTIVES

Our project objectives are designed to build upon experience gained under prior Office of Naval Research (ONR) support. This annual report describes progress attained in projects led by PI Milliff in the first full year of funding. First year results were also presented at a project workshop held at the Courant Institute for Mathematical Sciences, New York University, in November 2011. Objectives addressed in this annual report focus on extensions of a time- and space-dependent vertical error covariance BHM from the Mediterranean Forecast System (MFS) to the Regional Ocean Model System (ROMS) applications in the California Current System (CCS).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**
4. TITLE AND SUBTITLE **Bayesian Hierarchical Model Characterization Of Model Error In Ocean Data Assimilation And Forecasts**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Colorado Research Associates Division, NWRA,3380 Mitchell Lane,Boulder,CO,80301**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

APPROACH

Time-Varying Error Covariance Models: Review

During the data assimilation step in an ocean forecast system, the forward model trajectory is adjusted by available observations. The degree of adjustment is a function of uncertainties in the ocean state vector estimates from the model and observational uncertainties. These uncertainties are quantified in Error Covariance matrices in the cost function minimization used to obtain an "optimal" adjustment of the model trajectory given the data. A typical cost function is given by:

$$J = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}[\mathbf{H}(\delta\mathbf{x}) - \mathbf{d}]^T\mathbf{R}^{-1}[\mathbf{H}(\delta\mathbf{x}) - \mathbf{d}]; \tag{1}$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ for state vector $\mathbf{x}$ and background state vector $\mathbf{x}_b$, $d = \mathbf{H}(\mathbf{x}) - y$ are misfits for observation operator $\mathbf{H}$ and observations y. The matrices $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices, respectively. The full background error covariance $\mathbf{B} = \mathbf{V}\mathbf{V}^T$ for a sequence of operators $\mathbf{V} = \mathbf{V}_D\mathbf{V}_{uv}\mathbf{V}_h\mathbf{V}_H\mathbf{V}_V$ for diffusion (D), nonlinear advection (uv), surface height (h), horizontal (H) and vertical (V) variability.

The ultimate object of a time-varying background error covariance BHM is a space-time model for $\mathbf{B}$. For purposes of our initial developments we focus on only the vertical variability. So for now let, $\mathbf{B} = \mathbf{V}_V\mathbf{V}_V^T$, and $\delta(\mathbf{x}) = [\delta\mathbf{T}(z), \delta\mathbf{S}(z)]^T$, for temperature and salinity profiles, $T(z)$ and $S(z)$.

As noted in Berliner et al. (2003), and further codified in Cressie and Wikle (2011), BHM construction can be partitioned into data stage, process model and parameter model distributions for the components on the righthand side of Bayes Theorem. Gibbs sampling algorithms are developed to estimate posterior distributions of interest given these components. Estimates of the posterior distribution for processes (e.g. e_t below) and parameters (e.g. $\mathbf{B}_t$ below) are obtained and analyzed.

Let $\mathbf{e}_t$ be a space-time variable ocean forecast model error process. Relying on previous experience with basis function expansions and random, time-dependent amplitude coefficients (e.g. Wikle et al. (2001); Milliff et al. (2011)), we pose the process model:

$$\mathbf{e}_t = \Phi\beta_t + \eta_t \tag{2}$$

where Φ are a truncated set of vertical EOF bases, β_t are time-dependent amplitudes, and $\eta_t \sim Gau(0, \sigma_\eta^2 I)$ account for additional uncertainty, such as that arising from the dimension reduction due to truncation of the basis function set. Critically, we assume that $\beta_t \sim Gau(0, \mathbf{B}_t)$, where $\mathbf{B}_t$ is the time-dependent background error covariance matrix of interest in the MFS context.

The data stage inputs to the error covariance BHM are *model misfits* $\mathbf{d}_t$ noted above and *anomalies* $\mathbf{q}_t$, where the anomalies are daily departures from the model "year-day" climatologies; i.e. $\mathbf{q}_t = \mathbf{x}_{t|t-1} - \bar{\mathbf{x}}$ for the state vector climatology $\bar{\mathbf{x}}_t$ for a given forecast model.

The original error covariance BHM developments, supported by ONR, were implemented in the MFS where $\mathbf{B}_t$ is the focus. Impact studies for the MFS implementation of $\mathbf{B}_t$ are in progress now and will be reported and published in the coming year. Current ONR funding supports the extension of the error covariance BHM to the CCS domain of the ROMS ocean forecast system implemented at UCSC. Here, our primary interest is in the error process model itself; i.e. $\mathbf{e}_t$ as defined in (2).

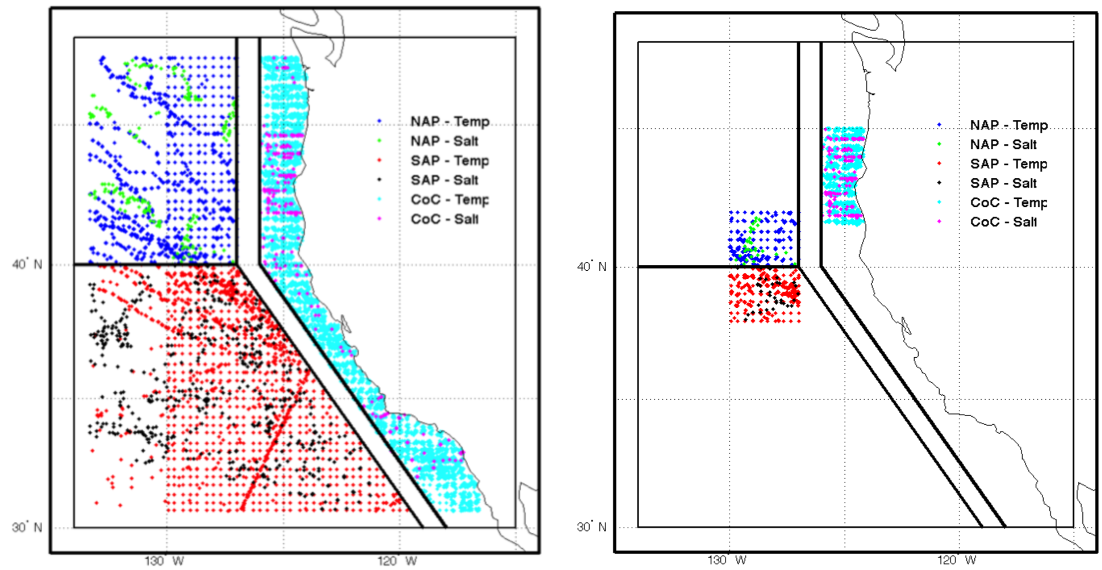

***Figure 1:** California Current System domain and data coverage by sub-region. Sub-regions include coastal (CoC), northern and southern abyssal plains (NAP, SAP). Data include temperature and salinity profiles. Subsets of the full dataset are shown at right. The data subsets have been used in the development and first implementations of the Error Covariance BHM in the CCS.*

WORK COMPLETED

The CCS domain is partitioned into 3 sub-regions as shown in Figure 1. The coastal domain (CoC) contains continental shelf regions. Offshore domains are separated in the vicinity of the Mendocino escarpment. They are denoted the Northern and Southern abyssal plain sub-regions (NAP and SAP respectively). Fig 1 also shows the distribution of temperature and salinity profiles for CCS by sub-regions. Subsets of these data (Fig. 1; right) were used in the first developments and implementations of the error covariance BHM in the CCS.

Figure 2 shows time vs. depth aggregates of the temperature profile data for the coastal (CoC; top) and N. abyssal plain (NAP; bottom) data subsets. The misfit data, **d** (left panels), are only available at measured profile locations and times, while the anomaly data **q** (right panels) are available for every year-day in the forecast period.

The first error process model estimates, for subregions of the CCS, are described below.

Relevant Meetings and Presentations

(Edwards, Milliff, Moore) Experimental design considerations, on-site visit with collaborators at Univ. California, Santa Cruz; Februrary, 2011.

(Herbei, Milliff, Wikle) Error covariance test case discussions, Advection-Diffusion model priors for Model Error, Herbei, Wikle visit to NWRA/CoRA; June 2010.

(Edwards, Milliff, Moore) Model Error in 4DVar discussions, on-site visit with collaborators at Univ. California, Santa Cruz; July, 2011.

(Berliner, Herbei, Milliff, Moore, Wikle) Informal presentations and discussions at the annual

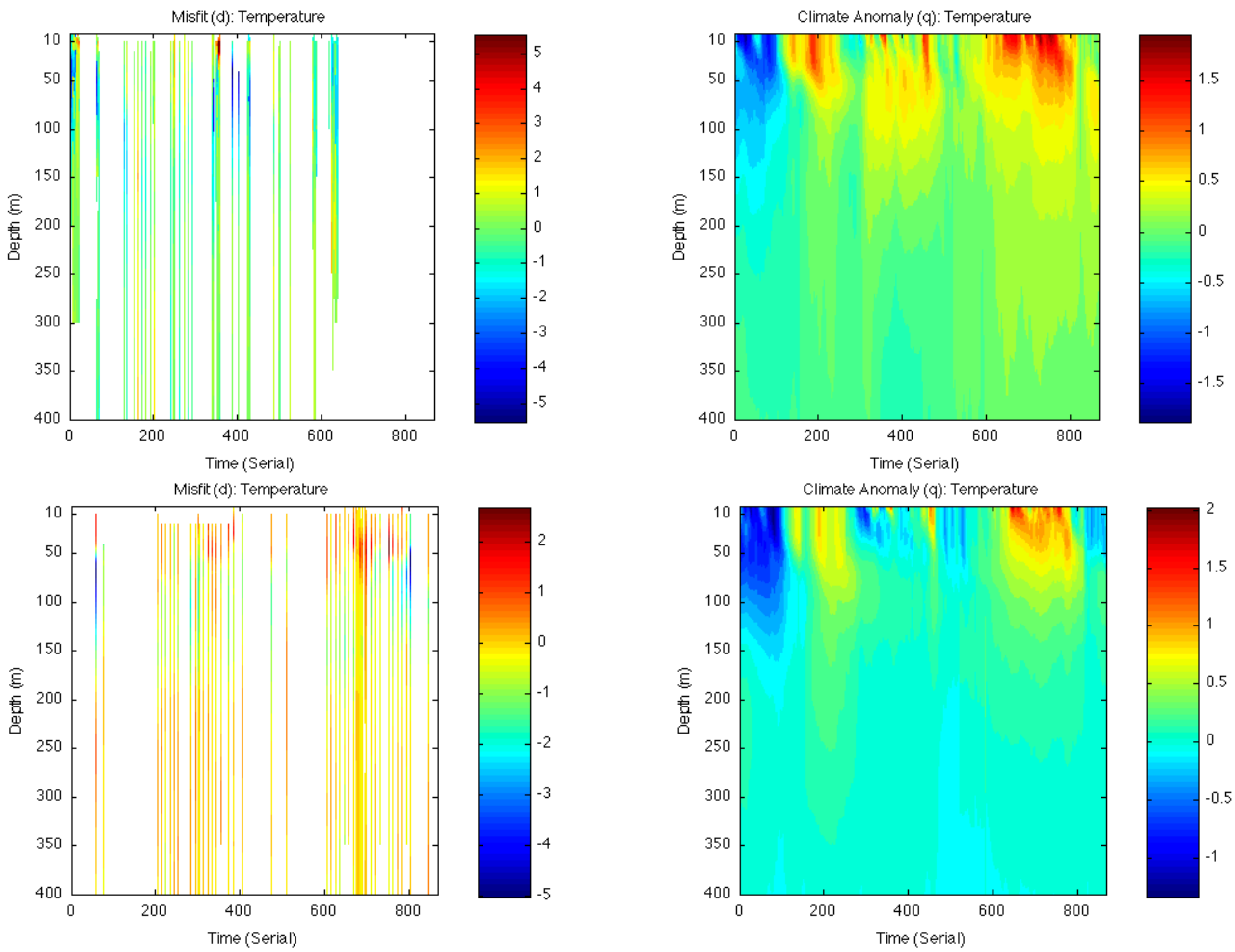

***Figure 2:** Temperature profile misfit (left panels) and anomaly (right panels) data stage inputs used for the development of a time-dependent error covariance BHM application in the California Current System region. The BHM data stage inputs are shown for the coastal (CoC; top) and N. abyssal plain (NAP; bottom) sub-domains. Data stage inputs for salinity are not shown.*

"All-Hands" project meeting at NWRA/CoRA, August, 2011.

(Milliff) Session leader; Stochastic forecasting and relation to Data Assimilation at, Guiding the Extension of Navy Operational Ocean Data Assimilation and Prediction, NRL SSC sponsored meeting at Univ. Maryland; September, 2011

(Herbei, Milliff, Moore) Presentations at ONR Model Error Project Meeting, New York Univ.; November, 2011.

RESULTS

Posterior mean time-dependent error, for errors manifest in temperature and salinity (not shown) profiles, are computed for the CCS domain of the ROMS 4dVar data assimilation and ocean forecast system. Data stage inputs from ROMS 4dVar misfits and the ROMS climatology for the region are convolved with basis function process models (2) as described above. Uncertainty in the error process is also time dependent and can be represented by time series of standard deviations in posterior distributions from the BHM.

Figure 3 depicts depth vs. time displays of the error process for temperature, $e_T(t)$ for the coastal and N. abyssal plain sub-regions of the CCS. Time series for standard deviation of the error process are shown at right (Fig. 3) as computed from the spread in the posterior distributions. The experimental domain focuses on the upper $400\,m$ in each sub-region; for a $175\,d$ period, treated in $5d$ epochs so that data stage influences can be maximized.

IMPACT/APPLICATIONS

The research overlapping the ONR project to use BHM to augment MFS, with the initial year of the ONR model error project demonstrates practical methods to add time- and space-dependence to error process and error covariance representations in operational (MFS) and near-operational (ROMS-4dVar) ocean forecast systems. Refining estimates of the time-dependent changes in forecast uncertainty across regime shifts adds value to ocean forecast system output.

TRANSITIONS

The Bayesian Confab meetings in Boulder every August are adding Irreducible Model Error foci in the informal presentations and discussions that characterize the meetings.

Informal communications with scientists in the Ocean Modelling branch of the Naval Research Laboratory, Bay St. Louis, MI have carried over from the ONR MFS project.

RELATED PROJECTS

"Bayesian Hierarchical Models to Augment the Mediterranean Forecast System", ONR Physical Oceanography Program, May 2009 - May 2011.

"Estimating Ecosystem Model Uncertainties in Pan-Regional Syntheses and Climate Change Impacts on Coastal Domains of the North Pacific Ocean", NSF US Globec Program, October 2008 - October 2011.

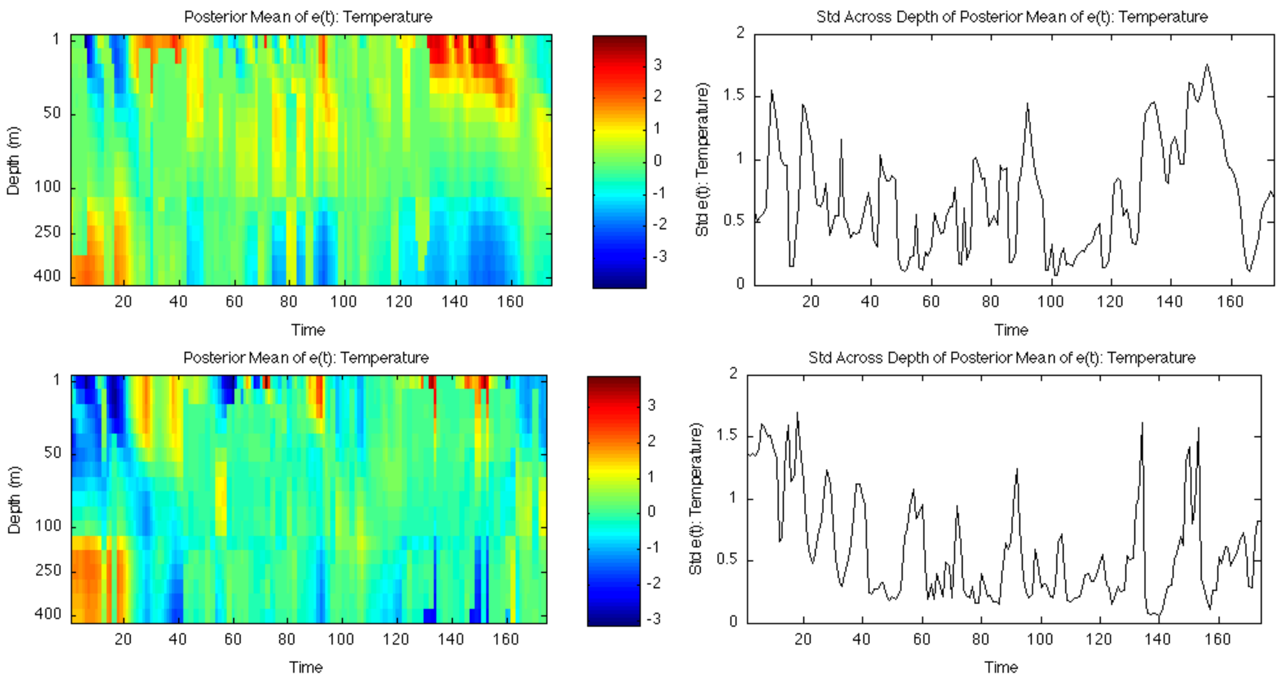

Figure 3: ***Posterior distribution temperature error process $e_T(t)$ (left panels), and temperature error process uncertainty (right panels) in a time vs. depth format for coastal (CoC; top) and N. abyssal plain (NAP; bottom) sub-regions. Error process uncertainty is expressed in standard deviation $e_T(t)$. Error process and uncertainty profiles are shown for a 175-day experiment spanning the period of greatest observational density in the data stage input records (Fig. 2). Units are $^{\circ}C$.***

"Quantifying the Amplitude, Structure and Influence of Model Error during Ocean Analysis and Forecast Cycles", ONR Physical Oceanography Program, A. Moore (PI).

REFERENCES

Berliner, L.M, R.F. Milliff and C.K. Wikle, 2003: "Bayesian hierarchical modelling of air-sea interaction", *J. Geophys. Res.*, **108**(C4), 3104, doi:10.1029/2002JC001413.

Cressie, N. and C.K. Wikle, 2011: **Statistics for Spatio-Temporal Data**, John Wiley and Sons Inc., 588 pgs.

Milliff, R.F., A. Bonazzi, C.K. Wikle, N. Pinardi and L.M. Berliner, 2011: "Ocean Ensemble Forecasting, Part I: Mediterranean Winds from a Bayesian Hierarchical Model", *Quarterly Journal of the Royal Meteorological Society*, **137**, 858-878.

Pinardi, N., A. Bonazzi, S. Dobricic, R.F. Milliff, C.K. Wikle and L.M. Berliner, 2011: "Ocean Ensemble Forecasting, Part II: Mediterranean Forecast System Response", *Quarterly Journal of the Royal Meteorological Society*, **137**, 879-893.

Wikle, C.K., R.F. Milliff, D. Nychka, and L.M. Berliner, 2001: Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *J. Amer. Statist. Assoc.*, **96**, 382-397.

PUBLICATIONS

The recent book by Cressie and Wikle (2011) and the *Quarterly Journal* papers by Milliff et al. (2011), and Pinardi et al. (2011), benefited from the continuing ONR support for BHM developments. See references.